DOCUMENT RESUME

ED 143 693 TM 006 458

AUTHOR Burstein, Leigh; Linn, Robert L.
TITLE The Identification of Teacher Effects in the Presence of Heterogeneous Within-Class Relations of Input to Outcome.
PUB DATE Apr 77
NOTE 32p.; Paper presented at the Annual Meeting of the American Educational Research Association (61st, New York, New York, April 4-8, 1977)

EDRS PRICE MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS *Academic Achievement; Classrooms; *Data Analysis; Educational Research; Mathematical Models; Measurement Techniques; *Multiple Regression Analysis; Research Methodology; *Research Problems; *Statistical Analysis; Students; *Teacher Influence

ABSTRACT

Multiple sources of complications in the analysis of multilevel educational data are described with emphasis on the determination of a method of analysis that will provide accurate estimates of teacher/class effects even when there are systematic differences in within-class regressions of outcome on input associated with teacher/class characteristics. Multilevel approaches are considered as well as traditional pupil-level and class-level analytical models. The results from a simulated model of educational effects which allowed the effects of teacher quality on the variation in the within-class slopes to vary over a range of meritocratic, random and compensatory conditions illustrated that overall between-student analyses, between-class analyses, and proposed multilevel methods can all yield misleading estimates of the magnitude of teacher effects on mean class outcome. However, selected multilevel methods provide some indication of misspecification and can identify the direction of the bias in estimating teacher effects on mean class outcomes. (Author/MV)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED143693

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

The Identification of Teacher Effects in the Presence of Heterogeneous Within-Class Relations of Input to Outcome[1]

Leigh Burstein
University of California, Los Angeles

Robert L. Linn
University of Illinois[2]

As part of a Panel Discussion on Analysis of Multilevel Data in Educational Research

PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL HAS BEEN GRANTED BY

Leigh Burstein

TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE NATIONAL INSTITUTE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER.

Presented at the Annual Meeting of the American Educational Research Association, New York, April 4-8, 1977.

[1]This paper was partially supported by NIE through the Center for the Study of Evaluation at UCLA. A more detailed version is available on request from the first author.

[2]The second author is presently a Visiting Professor at UCLA and Visiting Scholar at the Center for the Study of Evaluation.

TM006 458

ABSTRACT

Multiple sources of complications in the analysis of multilevel educational data are described and a single concern -- the determination of a method of analysis that will provide accurate estimates of teacher/class effects when there are systematic differences in within-class regressions of outcome on input that are associated with teacher/class characteristics -- is examined. Multilevel approaches are considered in addition to traditional pupil-level and class-level analytical models.

The results from a simulated model of educational effects which allowed the effects of teacher quality on the variation in the within-class slopes to vary over a range of meritocratic, random and compensatory conditions illustrated that overall between-student analyses, a between-class analyses, and proposed multilevel methods can all yield misleading estimates of the magnitude of teacher effects on mean class outcome. However, selected multilevel methods provide some indication of misspecification and can identify the direction of the bias in estimating teacher effects on mean class outcomes.

THE IDENTIFICATION OF TEACHER EFFECTS IN THE PRESENCE OF HETEROGENEOUS WITHIN-CLASS RELATIONS OF INPUT TO OUTCOME

Efforts to identify the effects of education (e.g., Coleman et al., 1966) on pupil performance have suffered from a lack of attention to the complications caused by the multilevel character of educational data. Schools are aggregates of their teachers, classrooms and pupils, and classrooms are aggregates of the persons and processes within them. This being the case, the effects of education exist in one form or another both between and within the units at each level of the educational system. Yet the majority of studies of educational effects has restricted attention to either overall between-student, between-class, or between-school analyses.

Cronbach (1976) argued that the majority of studies of educational effects carried out thus far conceal more than they reveal, and that "the established methods have generated false conclusions in many studies" (p. 1). His concern is foreshadowed in the educational literature by the exchange among Wiley, Bloom, and Glaser as recorded in Wittrock and Wiley (1970), and by Haney's (1974) review of the units of analysis problems encountered in the evaluation of Project Follow Through.

In this paper, we attempt to focus on a concrete manifestation of the complexities alluded to above. We are interested in illustrating selected analytical consequences when the data consist of scores associated with individual pupils who are nested within classrooms which are in turn nested within schools and the primary questions concern the effects of teachers on pupil performance with only secondary interest in the effects of more global school characteristics.[1] Research on classrooms such as that carried out by the Educational Testing Service (McDonald and Elias, 1976) and by the Far West Laboratory for Educational Research (Berliner, 1976) as

part of the Beginning Teacher Evaluation (BTES) Study attempt to answer this type of question.

Several alternative analytical strategies will be considered. These startegies include single level models (e.g., between-student analysis, between-class analysis and selected multilevel models (Burstein, 1976; Burstein and Linn, 1976; Cronbach, 1976; Cronbach and Webb, 1975; Keesling and Wiley, 1974). Strategies are characterized as multilevel if they require analysis in at least two stages for at least two levels of units. The differences in the analytical consequences from the various strategies are illustrated by comparisons using simulated data in which the magnitude (effect on mean performance) and form (effect on within-class slopes) of the effects of the teacher have been varied.

Issues in the Choice of Units of Analysis

Traditionally, a variety of competing points have been cited as justification for the choice of either pupils or groups (classrooms, schools, etc.) as the appropriate units of analysis in studies of educational effects. Both conceptual and statistical arguments have been voiced in favor of either level. A few of the key arguments for pupils or groups as units of analysis and selected relevant references are stated below.

(1) In education the phenomena we wish to investigate are pupil outcomes. More specifically, we want to determine the effects of the school (class, teacher) resources an individual pupil receives, his background and the influence of his community setting and peers on his educational outcomes (Averch et al., 1972; Burstein, 1975, 1976b; Burstein and Knapp, 1975; Burstein and Smith, 1977). Therefore, pupils are the units for which questions must finally be answered.

(2) Pupils react as individuals and the effects on them should be the focus of educational evaluation (Bloom in Wittrock and Wiley, 1970, p. 271ff.).

(3) The effects in classrooms are an aggregation of effects of environmental arrangements on individuals (Glaser in Wittrock and Wiley, 1970, pp. 271ff).

(4) Theoretical arguments concerning the effects of educational structure on pupil outcomes are formulated at the pupil level. To analyze the data at the group level is to enhance the likelihood of specification bias and aggregation bias (Hannan, Freeman and Meyer, 1976).

(5) The appropriate unit of study in educational evaluation is the collective -- class or school -- rather than the individual. The effects of a treatment on the classroom are fundamentally different from the effects of the treatment on the individuals within the classroom (Wiley in Wittrock and Wiley, 1970, pp. 271ff).

(6) The sampling unit determines the unit of analysis. If classrooms are the sampling units, then classrooms are the unit of analysis (Cline et al., 1974; Cronbach, 1976).

(7) The unit of treatment defines the level of analysis. If treatments are administered to intact classrooms, classrooms are the units (Cronbach, 1976; Glass and Stanley, 1970; Peckham et al., 1969).

(8) Pupil performances within classrooms are generally correlated with each other. These dependencies dictate the choice of between-class analyses (Glass and Stanley, 1970; Glendening and Porter, 1974; Glendening, 1976).

(9) Characteristics of the teacher take on the same value for every pupil in a particular classroom. An analysis at the pupil level overemphasizes the amount of information one has about class-level variables (Keesling and Wiley, 1974).

(10) The same manifest variables answer different questions at different levels of analyses. For example, a pupil's perception of classroom climate has a distinctly different meaning than the class aggregate perception of classroom climate. The sex of a student's teacher characterizes a social psychological effect, while the average teacher sex in a school characterizes an organizational process. Thus, the research foci should determine the appropriate unit of analysis (Burstein, 1976b; Cronbach, 1976; Scheuch, 1966).

(11) Dependence among observations within classrooms is a matter of degree rather than existence. Furthermore, the analysis of class means can mask between-class differences in the within-class distributions of outcomes and differences in the within-class regressions of outcomes on inputs. Thus, the use of class as the unit and means as the only class statistic used in the analysis is questionable (Brown and Saks, 1975; Burstein, 1976; Klitgaard, 1974, 1975; Lohnes, 1972).

(12) Overall between-student analyses are weighted averages of between-class and pooled within-class analyses and are thus rarely advisable in educational contexts (Cronbach, 1976).

The arguments cited above are compelling and virtually unresolvable if a choice of either pupil or class as the only unit of analysis is required. The key to a potentially viable solution is contained in points (11) and (12). The multilevel character of educational data warrants analytical strategies tailored to the identification of educational effects at and within each level of the educational system. The remaining discussion focuses on such analysis strategies and on one specific complication -- heterogeneous within-class slopes -- that may be encountered in their use.

Decomposing Pupil Outcomes and Teacher Effects

Once the existence of specific class membership is acknowledged

(i.e., instruction from a specific teacher)[2], any measure that varies over pupils can be decomposed into its between-class (teacher) and within-class (teacher) components. That is, the posttest or outcome performance, Y_{ij}, of pupil j in class i (j = 1, ..., n persons per class; i = 1, ..., k classes; for simplicity, we assume equal-size classes) can be decomposed into

$$(1) \quad \underset{\text{individual outcome}}{Y_{ij}} = \underset{\text{grand mean}}{\bar{Y}_{..}} + \underset{\text{between-class effect}}{(\bar{Y}_{i.} - \bar{Y}_{..})} + \underset{\text{within-class effect for person ij}}{(\bar{Y}_{ij} - \bar{Y}_{i.})}$$

If, in addition, we consider the performance level, X_{ij}, of the pupil prior to entering the class (i.e., the pretest or some measure of entering ability), then the relation of X_{ij} to Y_{ij} can also be decomposed into between-class and within-class components.

Following Cronbach (1976, pp. 3.1 - 3.11), this decomposition can be written as

$$(2) \quad Y_{ij} - \bar{Y}_{..} = \beta_b(\bar{X}_{.} - \bar{X}_{..}) \qquad \text{Predicted Between-Class Effect}$$

$$+ \bar{Y}_{i.} - \beta_b(\bar{X}_{i.} - \bar{X}_{..}) \qquad \text{Adjusted Between-Class Effect}$$

$$+ \beta_w(X_{ij} - \bar{X}_{i.}) \qquad \text{Pooled Within-Class Effect}$$

$$+(\beta_i - \beta_w)(\bar{X}_{ij} - \bar{X}_{i.}) \qquad \text{Specific Within-Class Effect}$$

$$+ \varepsilon_{ij} \qquad \text{Specific Residual Associated with Person ij}$$

In the above equation, β_b is the between-class slope from the regression of $\bar{Y}_{i.}$ on $\bar{X}_{i.}$, β_w is the pooled within-class slope from the regression of $(Y_{ij} - \bar{Y}_{i.})$ on $(X_{ij} - \bar{X}_{i.})$ <u>across all</u> classrooms, and the β_i are the <u>specific</u> within-class slopes from the regression of Y_{ij} on X_{ij} within the ith classroom.

The possible substantive interpretations of specific components and sets of components are important here. For the present discussion, the key

elements are the between-class slope, the adjusted within-class effect, the pooled within-class slope, and the specific within-class slopes.

Between-Class Effect. If positive, as it typically would be, the between-class effect, β_b, reflects the tendency of classes with higher mean inputs to also have higher mean outcomes. A set of classrooms characterized by high β_b may have experienced substantial or negligible average gains throughout the range of inputs. Relative to another set of classrooms with similar inputs and average gains but a low β_b, however, the set with the high β_b has been meritocratic rather than compensatory. That is, either groups at all levels of input have benefited by roughly equal amounts (i.e., constant gain on the scale of the measuring device) or by an amount directly proportional to their initial level. Conversely, low values of β_b reflect either redistribution of educational performance in a manner which bears little relationship to average entering characteristics or perhaps simply the irrelevance of average entering characteristics to outcomes. In either case, the relative mean outcome performance of classrooms with low mean entering performance will be higher for a set of classrooms characterized by low β_b than for a set of classrooms characterized by high β_b assuming that there is no average difference for the two sets.

Adjusted Between-Class Effect. The adjusted between-class effects $[\bar{Y}_{i.} - \beta_b(\bar{X}_{i.} - \bar{X}_{..})]$ reflect the effect of the specific teacher/class on the mean inputs on mean outcomes. Included in adjusted class effects are the main effect for the teacher, plus such things as unusual group cohesiveness or disharmony, high or low quality program delivery (e.g., well-structured, individualized instruction) and average error of measurement of Y_{ij} for the class. Typically, errors of measurement are not constant throughout the range of the measuring device and classes whose members fall at the extremes of the continuum may tend to have non-zero average errors of measurement. Also variability across classes in the conditions for

administering a measuring device (e.g., a teacher giving his/her pupils the answers) can result in the average measurement error.

In any event, large positive adjusted classroom effects may be educationally desirable. Whatever analysis strategy is employed should generate accurate estimates of this teacher/class effect on mean class outcome and, hopefully, should identify generalizable characteristics of teachers/classes achieving large effects.

Pooled Within-Class Effects. The pooled or common within-class effect, β_w reflects the tendency across all classes of students above the class average on input to do better or worse on the outcome than the rest of the class. The interpretation of β_w parallels that of β_b where the former refers to the consistent tendencies of the process within class, while the latter deals with tendencies of class averages. The usual educational finding is that classes tend to be strongly meritocratic (high positive β_w)[3] i.e., students with higher inputs gain more from a given increment of instruction than students with lower inputs. An alternative compensatory or redistributive model would have teachers bringing up pupils with lower inputs at a faster rate than they increase the performances of higher ability pupils. If teachers had relatively greater compensatory effects, β_w would still be positive, most likely, but closer to zero than in meritocratic classes.

Estimation of the common within-class regression is also of value in the study of teacher effects. β_w provides an indication of the overall redistributive properties of the classroom instruction encountered in the school population represented in the study. If policy makers and teacher educators are intent on raising the relative performance of low ability and low SES youth through better teaching methods and more skilled teachers, β_w indicates the magnitude of the problem they will encounter.

Specific Within-Class Effects. The within-class regression of pupil outcome on pupil inputs (which we denoted by β_i) is likely to var across

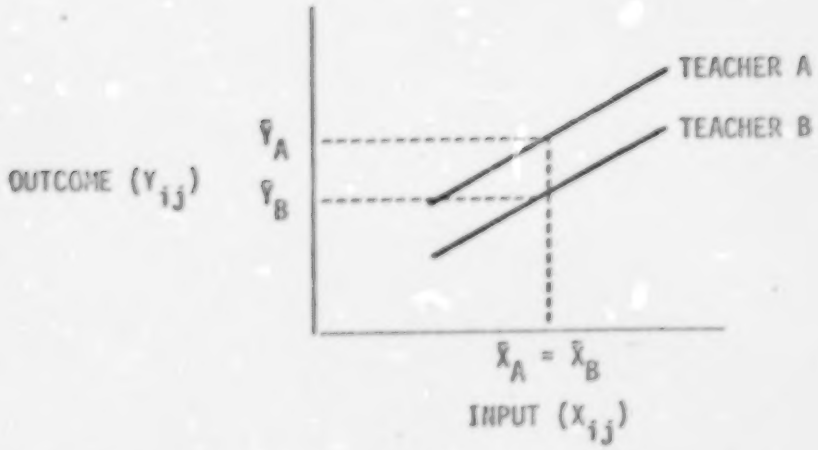
classrooms. Cronbach (1976, p. 316) cites as sources of the variation in β_i: (a) sampling variability due to chance and stability problems due to small class sizes when the processes operating in the classes are the same, (b) differences in the selection factors forming the classes, and (c) differences in casual processes going on in the classrooms. This last source encompasses the possibility that some teachers have relatively greater meritocratic effects on pupil outcomes while others have relatively greater compensatory (redistributive) effects or simply that teachers differ in the degree to which they are meritocratic.

If we could rule out stability problems (a big if) and different selection rules as reasonable explanations, the variation in β_i would become a potent source of information to researchers and policy makers, especially when such information is combined with the adjusted class effects discussed earlier. One can argue that isolation of teacher attributes and skills that are associated with compensatory or meritocratic within-class effects offers a cost-benefit in terms of the match of pupils, teachers, and schools that cannot be obtained by the consideration of adjusted class effects on mean outcomes.

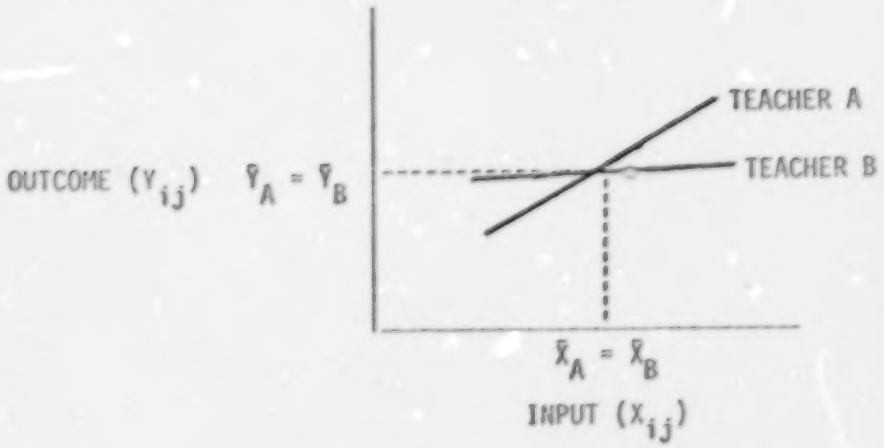
Actually, we can view a specific teacher's effect on the overall performance level of the class and on the relative performance of pupils within the class in a variety of ways. Several alternatives are illustrated in Figure 1. In each case, we have assumed that input and outcome scores are available from the pupils in two classrooms of equal size and with the same distribution of entering performance.

The three panels reflect mean effects only (panel (a)), slope effects only (panel (b)) and combined mean and slope effects (panel (c)), respectively.

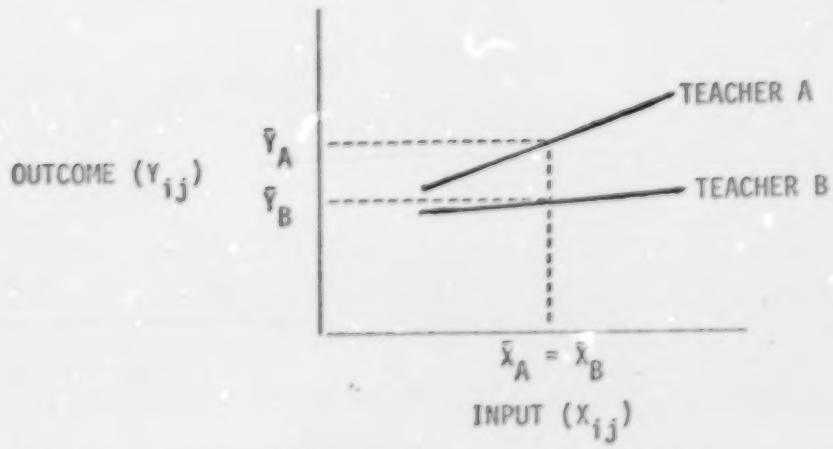
Insert Figure 1



(a) Teacher Effect on Mean Outcome; No Teacher Effect on Slope



(b) No Teacher Effect on Mean Outcome; Teacher Effect on Slope



(c) Teacher Effect on Mean Outcome; Teacher Effect on Slope

Figure 1. Hypothetical Teacher Effects: Same Range of Entry

A key question that should be asked about analysis strategies employed with classroom data is whether they are sensitive to the teacher effects depicted in Figure 1. For example, a between-class analysis of class means (either analysis of variance with classes as units, and teacher quality as the independent variable, or analysis of covariance with input as the covariate) would accurately identify any teacher quality effects on mean performance in all three panels. However, an analysis of class means will not distinguish the effects depicted in panel (a) (effect on mean outcome only) from the effects in panel (c) (effect on mean outcome and slope). Nor would such an analysis detect that teacher (A) and teacher (B) from panel (b) have different effects on pupils at different levels of input even though the mean performance of their pupils is the same.

Figure 2 depicts teacher effects on class mean and/or on slopes when the classes also differ in the range of entering performance. Note that in panels (b) and (c), the component for the between-class effect (β_b) from equation 2 had to be introduced in order to operationalize the concept of no teacher effect on class means for a comparison of two teachers. By saying that there are no differences in the effects of teacher A and teacher B on class mean performance (panel (b)), we mean that the two teachers have the same adjusted class effects:

$$\bar{Y}_A - \beta_b(\bar{X}_A - \bar{X}_{..}) = \bar{Y}_B - \beta_b(\bar{X}_b - \bar{X}_{..}) \quad .$$

(see second component of equation (2)).

Insert Figure 2

We are again interested in the sensitivity of alternative analysis strategies to the depicted teacher effects. The results of a between-class analysis of class means will differ in one important respect from the results for the cases depicted in Figure 1. Though accurate estimates of teacher effects on mean class performance are still possible for the

cases in panels (a) and (b), the magnitude of the teacher effect on mean performance is overestimated for the case depicted in (c). And, as with the cases depicted in Figure 1, an analysis of means alone would obviously fail to detect between-class differences in slope for the cases depicted in Figure 2.

Ideally, the researcher would want to control for slope differences in assessing mean effects and control for mean differences in assessing slope effects. Theoretically, with robust estimation of specific within-class slopes, between-class regressions of class means and class slopes on teacher characteristics should enable the investigator to determine whether he is in case (a), (b) or (c) from Figure 1. Our simulation below demonstrates how we expect this analytical strategy to operate.

We do not yet know the importance of differences in specific within-class regressions. Regressions based on populations of size 30 (classes are fixed populations ignoring transiency and absenteeism) are not highly robust with respect to outliers. A few atypical cases can dominate between-class differences in slopes (see later comments on Cronbach). The investigator obviously must eliminate outlier effects as a plausible explanation for heterogeneous slopes before proceeding with detailed analysis of the antecedents of slope differences. For the remainder of this paper, we will assume that the heterogeneous specific within-class slopes are not simply artifacts due to outliers, but instead represent a teacher/class characteristic when they occur.

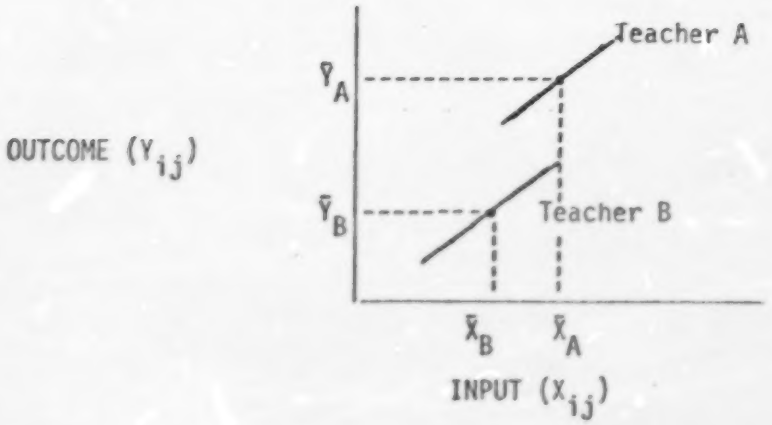
Alternative Analytical Models

Y_{ij} = Outcome of person j in class i;

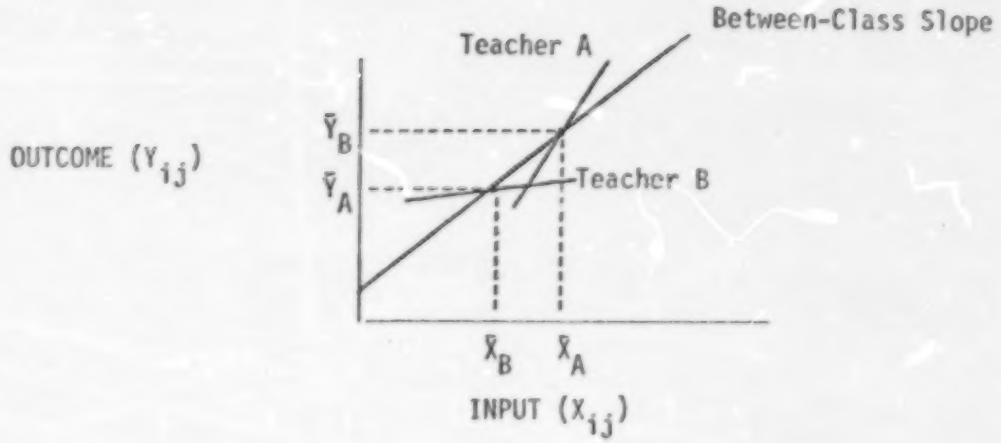
X_{ij} = Entering performance of person j in class i;

T_i = Teacher/class quality of class i;

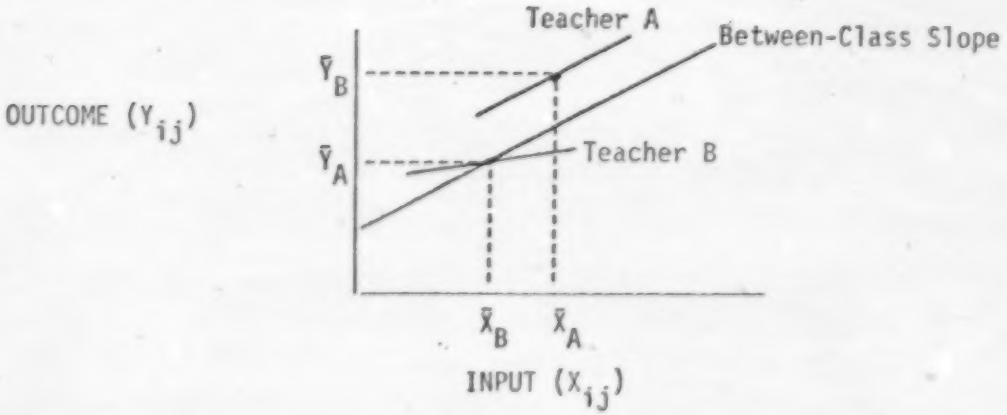
u_{ij}, v_i = Disturbance terms.



(a) Teacher Effect on Mean Outcome; No Teacher Effect on Slope.



(b) No Teacher Effect on Adjusted Mean Outcome; Teacher Effect on Slope.



(c) Teacher Effect on Adjusted Mean Outcome; Teacher Effect on Slope

Figure 2. Hypothetical Teacher Effects: Different Range of Entry

Between-Student Analysis

(3) $$Y_{ij} - Y_{..} = b_1(X_{ij} - \bar{X}_{..}) + b_2(T_{i.} - \bar{T}_{..}) + u_{ij}$$

b_1 = pooled within-class slope;

b_2 = adjusted class effect.

Between-Class Analysis

(4) $$\bar{Y}_{i.} - \bar{Y} = b_3(\bar{X}_{i.} - \bar{X}_{..}) + b_4(T_{i.} - \bar{T}_{..}) + \bar{v}_i$$

b_3 = between-class slope;

b_4 = adjusted between-class effect.

Between-Group, Pooled Within Group Analysis -- Cronbach (1976)

(6) $$\bar{Y}_{i.} - \bar{Y}_{..} = \underset{\sim}{b}_{\bar{Y}|T}T + \beta_b(\bar{X}_{i.} - \bar{X}_{..}) + \bar{v}_{i.}$$

Note: $\beta_b = b_3$

$b_{\bar{Y}T} = b_4$ when T is a continuous single variable.

(7) $$Y_{ij} - \bar{Y}_{i.} = \beta_w(X_{ij} - \bar{X}_{i.}) + (u_{ij} - \bar{u}_{i.})$$

Note: $\beta_w = b_1$ when T represents between-class differences.

Regression Analysis for Hierarchical Data -- Keesling and Wiley (1974)

Theoretical

(8) $$Y_{ij} - \bar{Y}_{i.} = \beta_w(X_{ij} - \bar{X}_{i.}) + \underset{\sim}{\gamma}\underset{\sim}{Z}_i + \underset{\sim}{\beta}_i(\underset{\sim}{Z}_i X_{ij}) + \underset{\sim}{\varepsilon}_{ij}$$

$\underset{\sim}{Z}_i$ = dummy variables for classes

$\underset{\sim}{\beta}_i$ = effects of class-pretest interaction

Practice

(9) $$\bar{Y}_{i.} = \underset{\sim}{\gamma}_0 + \underset{\sim}{\gamma}\bar{T} + \lambda(\hat{\bar{Y}}_i) + \phi_i$$

$\hat{\bar{Y}}_i$ = mean predicted outcome computed from aggregation of individual outcomes predicted from pooled within-group regression.

γ = effects of class-level variables denoted by $\underline{T}$, analogous to adjusted class effects.

λ = effects of predicted class mean (removes specification bias due to omission of relevant class-level variables).

Analyses of Slopes and Intercepts - Burstein (1976 and Burstein and Linn (1976)

$$(10) \quad \bar{Y}_{i.} - \bar{Y}_{..} = \underline{b}_{\bar{Y}T}\underline{T} + T_1$$

$$\beta_i = \underline{b}_{\beta T}\underline{T} + T_2$$

where

β_i = specific within-class regressions.

Simulation of Educational Effects

In an effort to determine the degree of commonality of results among the strategies described above, we simulated the following structural model[7]:

$$(11) \quad Y_{ijk} = b_{YT}T_{ij} + F_{ijk} + b_{ij}X_{ijk} + U_{1ijk}$$

$$T_{ij} = .1S_i + \sqrt{1 - (.1)^2}\, U_{2ijk}$$

$$S_i = .2F_{i..} + \sqrt{1 - (.2)^2}\, U_{3ijk}$$

$$F_{ijk} = (1/\sqrt{5})F_{i..} + (2/\sqrt{5})U_{4ijk}$$

$$X_{ijk} = .4F_{ijk} + \sqrt{1 - (.4)^2}\, U_{5ijk}$$

$$b_{ij} = 1 + c(.1)U_{2ijk} + \sqrt{1 - c^2}\, U_{6ijk}$$

One substantive interpretation of the model is that posttest performance for student k in class ij, Y_{ijk}, is a function of the student's entering

performance, X_{ijk}, the student's family background, F_{ijk}, the quality of instruction received, T_{ij}, and a random disturbance term, $U_{1ijk}(\sim N(0,1))$. Furthermore, teacher quality (T_{ij}) is a function of school quality (S_i)[8] which is, in turn, partly determined by the mean family background of the pupils attending the school ($\bar{F}_{i..}$). Also, the individual pupil's family background (F_{ijk}) affects his or her pretest performance. Finally, the U's, the disturbance terms inthe model, are all pseudo random numbers from independently distributed normal distributions with means and variances chosen for convenience.

The two model parameters, b_{YT} and c, were varied to reflect different degrees of dependence of posttest on teacher quality. b_{YT} reflects the direct effects of teacher quality on mean performance of the class. The values of $b_{YT} = 0$ (no direct effect) and $b_{YT} = .3$ (moderate effect) were chosen for the simulation.

The unique feature of the simulation is that the b_{ij}, the specific within-class regressions of outcome in input, vary across classrooms either randomly or as a systematic function of teacher quality. The parameter c determines the strength and the direction of the relationship of within-class slope to teacher quality. Positive c values cause higher quality teachers to have steeper slopes (be more meritocratic). Negative c values cause higher quality teachers to have flatter slopes (be more compensatory).

The analyses and discussion that follow are based on the generation of a single set of pseudo normal deviates to simulate a sample of $n_{ij} = 90$ classrooms with $n_k = 30$ pupils per classroom ($N = n_{ij}n_k = 2700$). Ten distinct sets of data were formed by varying the direct effect of teacher quality ($b_{YT} = .0, .3$) and the effects of teacher quality on the variation in b_{ij} ($c = .8, .4, .0, -.4, -.8$).

Single-Level Analysis. The pupil-level and class-level regressions of outcome on pretest, family background, and teacher quality (Tables 3 and 4) provide further support for our concern about systematic class-to-class variations in slopes. The estimated partial regression coefficients for the regression of Y on T, F and X are reported in Table 3 for the between-student analysis and in Table 4 for the between-class analyses. The estimates are reported for the various combinations of the simulation parameters (β_{YT} and c). All estimates have been adjusted to remove perturbations in the simulation such that $\hat{b}_{YT} = \beta_{YT}$ when c = 0. Under the assumption that we have identified the relevant antecedent variables (F, X, T) which determine pupil performance (a highly tenuous assumption which we consider later on), either between-student or between-class analyses yield the approxiamte parameter estimates as long as variation in within-class slopes is essentially random.

Tables 3 and 4

When teachers have either compensatory or meritocratic effects on within-class slopes, however, the picture becomes more complex. Though estimates of b_{YF} and b_{YX} are virtually unaffected in analyses at the pupil level, the estimates of b_{YT} are dramatically distorted in the direction of the effects of teacher quality on the slope. Moreover, in a between-class analysis, systematic variation of within-class slopes affects the estimation of both $b_{\bar{Y}\bar{X}}$ and $b_{\bar{Y}T}$.

It is important to keep in mind that we are focussing on the potential analytical consequences of heterogeneous within-class regressions in educational effects studies. Under such circumstances, models which assume

Table 4. Between-class regressions of outcome ($\bar{Y}$) on teacher quality ($\bar{T}$), family background ($\bar{F}$) and mean pretest ($\bar{X}$) as a function of the relation of teacher quality to outcome ($\beta_{\bar{Y}\bar{T}}$) and to variation in within-class slopes (c).

Simulation Parameters		Parameter Estimates		
$\beta_{\bar{Y}\bar{T}}$	c	$\hat{b}_{\bar{Y}\bar{T}}$	$\hat{b}_{\bar{Y}\bar{F}}$	$\hat{b}_{\bar{Y}\bar{X}}$
	.8	.37	.17	.87
	.4	.15	.20	.78
.0	.0	-.06	.22	.75
	-.4	-.26	.25	.75
	-.8	-.46	.27	.82
	.8	.67	.17	.87
	.4	.45	.20	.78
.3	.0	.24	.22	.74
	-.4	.04	.25	.75
	-.8	-.16	.27	.82

Table 3. Between-student regressions of outcome (Y) on teacher quality (T), family background (F) and pretest (X) as a function of the relations of teacher quality to outcome (β_{YT}) and to variation in within-class slopes (c).

Simulation Parameters		Parameter Estimates		
β_{YT}	c	$\hat{b}_{YT}$	$\hat{b}_{YF}$[a]	$\hat{b}_{YX}$[a]
	.8	.42[b]	.20	1.00
	.4	.21	.20	1.00
.0	.0	.00	.20	1.00
	-.4	-.20	.21	1.00
	-.8	-.38	.21	.97
	.8	.73	.20	1.00
	.4	.51	.20	1.00
.3	.0	.30	.20	1.00
	-.4	.10	.21	.98
	-.8	-.09	.22	1.00

[a]The parameters in the simulation are $b_{YF} = .2$, $b_{YX} = 1.0$.

[b]All estimates have been adjusted to remove effects due to perturbations in the simulation. The adjustments causes the values in the cases when $\beta_{YT} = .0$ or .3 and c = 0 to equal the specified parameter values. The magnitudes of the adjustments are .06 (T) .03 (F) and .01 (X) when $\beta_{YT} = .0$ and .07 (T), .02 (F) and .02 (X) when $\beta_{YT} = .3$.

a common within-class slope are misspecified and can be expected to result in biased estimation. So far, our main conclusion from the simulation is that neither pupil-level nor class-level analyses will yield correct estimates of teacher/class effects when there are systematic differences in within-class slopes that are determined by teacher quality.

Multilevel Analyses. Tables 5 and 6 present the results from analyzing the 10 sets of simulated data according to either the Keesling-Wiley (Table 5) or the slope-intercept strategy (Table 6).

As expected, the estimates of the effects of T on mean outcome $\bar{Y}$ across all simulated conditions are essentially the same for the Keesling-Wiley analysis (γ_T in Table 5), the slope-intercept analysis ($\hat{b}_{\bar{Y}T}$ in Table 6) and the two single-level analysis ($\hat{b}_{YT}$ in Table 3 and $\hat{b}_{\bar{Y}T}$ in Table 4). This occurs because we have assumed that we know the true model and that we can measure teacher quality on a single scale. In actual practice, neither of these assumptions is met in general. Nonetheless, all of the analytical strategies reflect the distortions in the estimates of direct teacher effects on pupil outcomes when there is systematic variability in within-class slopes which is related to teacher quality. All four strategies (and Cronbach's approach as well) will generate overestimates of the direct effects of teacher quality when the better teachers have steeper slopes (c = .4, .8) and underestimates of direct teacher effects when the better teachers have flatter slopes (c = -.4, -.8).

Tables 5 and 6

The potential advantage of either the slope-intercept analysis or the Keesling-Wiley analysis when compared to the between-student and the between-class regressions is derived from the additional information provided about the adequacy of the analysis of class means. The deviations of $\hat{\gamma}_{\bar{Y}}$ from

Table 5. Estimates of teacher effects ($\hat{\gamma}_T$) and model misspecification ($\hat{\gamma}_{\hat{Y}}$) from a Keesling-Wiley analysis.

Simulation Parameters		Parameter Estimates	
		Teacher Effect	Model Misspecification
β_{YT}	c	($\hat{\gamma}_T$)	($\hat{\gamma}_{\hat{Y}}$)[a]
	.8	.43[b]	1.03
	.4	.21	.99
.0	.0	.00	1.00
	-.4	-.20	1.05
	-.8	-.39	1.15
	.8	.73	1.03
	.4	.51	.99
.3	.0	.30	1.00
	-.4	.10	1.05
	-.8	-.09	1.15

[a]For a correctly specified class-level model, $\hat{\gamma}_{\hat{Y}}$ is expected to equal 1.00.

[b]Estimates of $\hat{\gamma}_T$ have been adjusted to remove effects due to perturbations in the simulation. The adjustment causes the values of $\hat{\gamma}_T$ to equal the specified parameter values when c = 0 and β_{YT} = 0 or .3 .

Table 6. Estimates of effects of teacher quality on class mean outcome ($b_{\bar{Y}T}$) and within-class slope ($b_{\beta T}$) from a slope-intercept analysis.

Simulation Parameters		Parameter Estimates	
β_{YT}	c	$\hat{b}_{\bar{Y}T}$	$\hat{b}_{\beta T}$
	.8	.44[a]	.09
	.4	.22	.05
.0	.0	.00	.00
	-.4	-.20	-.04
	-.8	-.40	-.09
	.8	.74	.09
	.4	.52	.05
.3	.0	.30	.00
	-.4	.10	-.04
	-.8	-.10	-.09

[a]Estimates of $\hat{b}_{\bar{Y}T}$ have been adjusted to remove effects due to perturbations in the simulation. The adjustment causes the values of $\hat{b}_{\bar{Y}T}$ to equal the specified parameter values when $c = 0$ and $\beta_{YT} = .0$ or $.3$.

1.0 suggest, perhaps, that the Keesling-Wiley strategy is sensitive to the problem of systematic variation in within-class slopes though apparently such effects must be of substantial magnitude to be detectable. If so, in the presence of specification errors such as the exclusion of a relevant variable and the existence of large systematic differences in within-class slopes, the Keesling-Wiley analysis sufficiently warns the investigator that caution is necessary in making inferences. The search for additional causes and the reconsideration of the adequacy of the pooled within-class adjustment are natural next steps for the forewarned investigator.

In contrast to the Keesling-Wiley analysis and the regressions at the pupil and class levels, the investigator can learn about the direction in which estimates of teacher effects are biased by examining $\hat{b}_{\beta T}$ from a slope-intercept analysis. For every combination of systematic slope differences in the simulation, $b_{\bar{Y}T}$ deviates from the β_{YT} in random variation case in the direction of the effect of teacher quality on within-class slopes. When c = .4 or .8, $b_{\bar{Y}T}$ overestimates β_{YT} regardless of level (.0 or .3) used in the simulation and when c = -.4 or -.8, $b_{\bar{Y}T}$ underestimates β_{YT}. The values .4, .0, -.4, and -.8 are the direct effects of T on variation in β_i. Since the standard deviation of β_1 in the simulation model was set at .1, the magnitudes of the effects can be expected to be .08, .04, .00, -.04 and -.08 which compare favorably with the estimated values of $\hat{b}_{\beta T}$ reported in the last column of Table 6. Thus, as can be seen, $\hat{b}_{\beta T}$ successfully captures the magnitude and direction of the effects of T on variation in β_i even when the effects are small.

The results from the slope-intercept analysis are not uniformly positive however. Unfortunately, at the two moderate values of c (.4, -.4), the effects of T on β_i might be considered nonsignificant by a standard hypothesis test. Nonetheless, the interpretability of teacher

effects is obviously enhanced by the examination of specific within-class slopes under the conditions considered here.

Conclusions and Caveats

Though multiple sources of complications in the analysis of multilevel educational data were cited, we have focussed on a single concern -- the determination of a method of analysis that will provide accurate estimates of teacher/class effects when there are systematic differences in within-class regressions of outcome on input that are associated with teacher/class characteristics. Multilevel approaches suggested by Cronbach (1976; Cronbach and Webb, 1975), Keesling and Wiley (1974), and Burstein (1976b; Burstein and Linn, 1976) were considered in addition to traditional pupil-level and class-level analytical models.

A model of educational effects was simulated wherein the effects of teacher quality on the variation in the within-class slopes was varied over a range of meritocratic (positive relation of teacher quality to within-class slope), random and compensatory (negative relation of teacher quality to within-class slope) effects. The direct effects of teacher quality on pupil outcome was also varied. Between-student and between-class regressions of outcome on all variables specified in the model were run in addition to the Keesling-Wiley and slope-intercept analyses. For the conditions reflected in the simulation, the following conclusions were reached regarding the identification of educational effects in the presence of systematic differences in between-class slopes:

(1) An overall between-student analysis, a between-class analysis, a Keesling-Wiley analysis and a slope-intercept analysis all yielded misleading estimates of the magnitude of teacher effects on mean class outcome.

(2) The Keesling-Wiley method of analysis provides some indication of misspecification due to deletion of relevant causes and perhaps due to systematic differences in within-class slopes.

(3) The slope-intercept method of analysis provides an indication of the direction of the bias in estimating teacher effects on mean class outcomes and may also suggest the severity of such biases.

The generality of our conclusions is limited by the restrictive nature of the simulation we carried out. We considered a relatively straightforward educational effects model generated from a single set of normal deviates whose only significant educative feature was the introduction of systematic teacher effects on within-class slopes. Moreover, the systematic effects we introduced were of the simplist form possible -- direct linear relations of teacher quality to variation in the within-class slopes. It is unlikely that actual differences in within-class slopes would arise in this straightforward fashion. Any judgments about the likely occurrence of biases as large or larger than those from the simulation when differences among within-class slopes are attributable to more complex sets of systematic factors would be purely speculative.

Heterogeneity of within-class slopes can have serious effects and thus deserves consideration whether by the procedures used above or some other approach. For example, a standard test of parallelism of regression slopes as in the analysis of covariance might be used to provide the same information generated by the Keesling-Wiley and slope-intercept analyses. Once heterogeneity of the within-class regressions is indicated, an examination of plots of within-class slopes versus variables like teacher quality (which may be unmeasurable) can lead to partial disentanglement of outliers. The investigator might also consider a model at the pupil level which examines the effects of teacher quality controlling for differences in

within-class slopes. Hannan, Rogosa and Young (personal communication) have suggested such a strategy as a viable alternative for overcoming the heterogeneity of regression problems.

In an earlier paper (Burstein, 1976) we argued that the development of approaches for the analysis of multilevel data is essential if we are to avoid looking at effects, one class at a time. Cronbach (1976) has stated this same concern more eloquently and with more caution about the possibility of developing a universally successful strategy. His caution is well justified if the results from our simulation of a fairly restrictive set of complications provide any indication of the difficulties that will be encountered in realistic settings.

FOOTNOTES

1. With no loss of generality, we could have chosen to consider the effects of schools on pupils rather than the effects of teachers. Studies such as the Coleman Report (Coleman et al, 1966) and the IEA Six Subjects Survey (e.g., Comber and Keeves, 1973),which employ primarily global school characteristics (e.g., availability of laboratory equipment, principal's experienced and school-level aggregates of teacher characteristics; e.g., average teacher experience, average teacher age) have essentialy two levels of analysis (pupil; school). However, such studies cannot identify the role of the pupil's own teacher on his or her educational performance.

2. For our purposes, we assume that class and teacher are completely confounded.

3. In practice an operationalization of the notion of meritocracy is called for. But if the value of β_w from a regression involving standardized variables approaches the typical value for pretest-posttest correlations (.7–.9), the process is surely more meritocratic than compensatory.

4. There is no loss in translating this method to a setting with pupils and classrooms as meaningful levels in the discussion that follows.

5. Their assumption of the non-existence of such interactions seems to be made more on the basis of complications in the computational algorithm than on theoretical grounds.

6. The slope-intercept method described here represents a modification of the procedures proposed in Burstein (1976b). We have dropped, for the time being, the idea of treating the standard error as an additional outcome variable and have used $\bar{Y}_i$, rather than α_i, as an outcome variable. The latter substitution was made to reduce the influence of the strong negative correlation between α_i and β_i across classrooms.

7. The specifics of the simulation are more complex than are presented here. We provide what we believe to be sufficient information to identify the emphasis and the implications of the simulated illustration

of model differences. Further details about the actual simulation are available from the authors.

8. We note in passing that our model posits no direct effects of school quality on pupil outcomes. In the simulation, we perceived school quality to influence performance through the ability to attract quality teaching or to organize classrooms in a manner conducive for high performance. The latter effect would also appear to be a teacher effect.

REFERENCES

Averch, H., Carroll, S.J., Donaldson, T.S., Kiesling, H.J. and Pincus, J. How Effective is Schooling?, The Rand Corporation, R-956-PCSF/RC, 1972.

Berliner, D.C., et al. Proposal for Phase IIIB of the Beginning Teacher Evaluation Study. San Francisco: Far West Laboratory for Educational Research and Development, 1976.

Brown, B.W. and Saks, D.H. "Proper Data Aggregation for Economic Analysis of School Effectiveness." Public Data Use, 1975, 3: 13-18.

Burstein, L. "The Choice of Unit of Analysis in the Investigation of School Effects: IEA in New Zealand." New Zealand Journal of Educational Studies, 1976a, 2: 11-24.

Burstein, L. "Assessing the Differences of Between-groups and Individual Regression Coefficients." Paper presented at the American Educational Research Association Annual Meeting, San Francisco, April, 1976b.

Burstein, L. "Data Aggregation in Educational Research: Applications." Paper presented at the Annual Meeting of the American Educational Research Association Annual Meeting, Washington, D.C., April, 1975.

Burstein, L. and Knapp, T.R. "The Unit of Analysis in Educational Research." Technical Report No. 4, Consortium on Methodology for Aggregating Data in Educational Research. Milwaukee, Wisc.: Vasquez Associates, Ltd., July, 1975.

Burstein, L. and Linn, R.L. "The Effects of Education in the Analysis of Multilevel Data: The Problem of Heterogeneous Within-class Regressions." Paper presented at a Conference on Methodology for Aggregating Data in Educational Research, Stanford, CA, October, 1976.

Burstein, L. and Smith, I.D. "Choosing the Appropriate Unit for Investigating School Effects." Australian Journal of Education, March, 1977 (in press).

Cline, M.G. et al. "Education as Experimentation: Evaluation of the Follow Through Planned Variation Model." Cambridge, Mass.: Abt Associates, 1974, Vols. 1A, (Ed 094890, 094891).

Coleman, J.S. et al. Equality of Educational Opportunity. Washington, D.C.: U.S. Department of Health, Education and Welfare, U.S. Office of Education, OE-38001, 1966.

Comber, L.C. and Keeves, J.P. Science Education in Nineteen Countries: An Empirical Study. International Studies in Evaluation, Volume I, Sweden: Almqvist & Wiksell, 1973.

Cronbach, L.J. "Research on Classrooms and Schools: Formulation of Questions, Design and Analysis." Occasional Papers of the Stanford Evaluation Consortium, July 1976.

Cronbach, L.J. and Webb, N. "Between-class and Within-class Effects in a Reported Aptitude Treatment Interaction: Reanalysis of a study by G.L. Anderson." Journal of Educational Psychology, 1975, 67: 717-727.

Glass, G.V. and Stanley, J.C. Statistical Methods in Education and Psychology. Englewood Cliffs, N.J.: Prentice-Hall, 1970.

Glendening, L. "The Effects of Correlated Units of Analysis: Choosing the Appropriate Unit." Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, April, 1976.

Glendening, L. and Porter, A.C. "The Effects of Correlated Units of Analysis: Violating the Assumption of Independence." A working paper: Measurement and Methodology Program, NIE, November, 1974.

Haney, W. "Units of Analysis Issues in the Evaluation of Project Follow Through." Unpublished Report, Cambridge, Mass.: Huron Institute, 1974.

Hannan, M.T., Freeman, J., and Meyer, J.W. "Specification of Models of Organizational Effectiveness." American Sociological Review, 1976, 41: 136-143.

Keesling, J.W. and Wiley, D.E. "Regression Models of Hierarchical Data." Paper presented at the Annual Meeting of the Psychometric Society, Stanford, CA, 1974.

Klitgaard, R.E. "Going Beyond the Mean in Educational Evaluation." Public Policy, 1975, 23: 59-79.

Klitgaard, R.E. Achievement Scores and Educational Objectives. Santa Monica, CA: The Rand Corporation, R-1217-NIE, 1974.

Lohnes, P.R. "Statistical Descriptors of School Classes." American Educational Research Journal, 1972, 9: 547-556.

McDonald, F.J. and Elias, P.J. The Effects of Teaching Performances on Pupil Learning, BTES Phase II Final Report, Educational Testing Service, 1976.

Peckham, P.D., Glass, G.V. and Hopkins, K.D. "The Experimental Unit in Statistical Analysis." Journal of Special Education, 1969, 3: 337-349.

Scheuch, E.K. Cross-national Comparisons Using Aggregate Data: Some Substantive and Methodological Problems. In R.L. Merritt and S. Rokkan (eds.), Comparing Nations: The Use of Quantitative Data in Cross-national Research, New Haven: Yale University Press, 1966, pp. 131-167.

Wittrock, M.D. and Wiley, D.E. The Evaluation of Instruction: Issues and Problems. New York: Holt, Rinehart & Winston, 1970.